REMARKS

In accordance with the foregoing, claims 34-43 have been cancelled, claims 1, 5, 15-19, 25, 29, 30 and 33 have been amended and claims 44-62 have been added. Claims 1-33 and 44-62 are pending and under consideration.

Independent claims 1, 29, 30 and 33 have been amended to convey that the tuples are reordered. Antecedent basis for these claim changes can be found, for example, at page 14, line 29 through page 15, line 7. Independent claims 50 and 51 have been added. These claims recite that filtering is based on parts of speech. Antecedent basis for these limitations can be found, for example, in the application at page 12, lines 22-25 and page 13, lines 1-16.

The Examiner is requested to note that the Office Action was sent to the previous attorney, at the Veneble law firm. As discussed several times with the Examiner, the attorney has changed. Applicants have received a notice indicating that the revocation/new power of attorney has been accepted. However, the notice received from the Patent Office indicates that Staas & Halsey is the former attorney. This is incorrect. The Examiner is requested to ensure that the undersigned is notified of any future action on this case. This is an important issue.

Turning now to the prior art rejections, claims 34-42 are rejected under 35 USC § 102(b) as being anticipated by Manber "Finding Similar Files in a Large File System," JSENIX, January 27 - 21, 1994. Although the rejected claims have been cancelled, Applicants will briefly discuss how the claims distinguish over the Manber reference. On page 1 of the Manber reference, it states "Files are considered similar if they have significant number of common pieces, even if they are very different otherwise." On page 2, the reference states "Our notion of similarity throughout this paper is completely syntactic. We make no effort to understand the contents of the files. Files containing similar information but using different words will not be considered similar. This approach is therefore very different from the approach taken in the information retrieval literature, and cannot be applied to discover semantic similarities. In a sense, this paper extends the work on approximate string matching . . ., expect that instead of matching strings to large texts, we match parts of large texts to other parts of large texts on a very large scale." On page 3, the reference states "We achieve the kind of synchronization described above with the use of what we call anchors. An anchor is simply a string of characters, and we will use a fixed set of anchors . . . The first is by analyzing text from many different files and selecting a fixed set of representative strings, which are quite common but not too common. The string *acte* is an example. Once we have a set of anchors, we scan the files we want to compare and search for all occurrences of all anchors. Fortunately, we can do it reasonably

quickly using our multiple-pattern matching algorithm . . . The second method computes fingerprints of essentially all possible substrings of a certain length and chooses a subset of these fingerprints based on their values."

It should be apparent that Manber does not reorder the document tokens to do the search. Manber uses strings. By definition, strings are in the order in which they appear. It should also be apparent that the Manber is not concerned with parts of speech. Manber states that it makes no effort to understand the contents of the files.

The Examiner also raises an obviousness rejection relying upon Brin et al. "Copy Detection Mechanisms for Digital Documents," ACM 1995, pages 398-409, in view of U.S. Patent No. 5,136,646 to Haber et al.

With regard to Brin et al, the Examiner should note that this reference was cited by Applicants in the specification. Brin et al. is an extension of the Manber reference. On page 400, column 1, Brin et al. states that "The resulting canonical form document consists of a string of ascii characters with whitespace separating words, punctuation separating sentences and possibly a standard method marking the beginning of paragraphs." This portion of the reference indicates that Brin et al. are purely concerned with syntax. The parts of speech are irrelevant in the Brin et al. inquiry.

At column 2 on page 400, Brin et al. states "We define a chunk as a sequence of consecutive units in a document of a given unit type. . . Then it can be organized into chunks as follows : A, B, C, D, E, F, G; or AB, CD, EF, G; or AB, BC, CD, DE, EF, FG; or ACB, CD, EFG; or A, D, G." It should be clear from the excerpts at column 2 on page 400 that Brin et al. employs no reordering, as certain claims require.

As to Haber et al., this reference is directed to a sophisticated, secure hash scheme. The digital document time stamping with certificate scheme stops people from changing time stamps. Haber et al. is simply directed to a computational technique. There is no disclosure regarding detecting similar documents, reordering documents or filtering documents.

On page 6 of the Office Action, the Examiner raises an obviousness rejection, relying upon Brin et al., Haber et al. and U.S. Patent No. 6,240,409 to Aiken. Aiken is also cited in the rejection beginning on page 4 (item 4), even though not specifically relied upon. Aiken describes at column 2, lines 36-41, "Therefore, it would be desirable to determine similarities among large sets of documents in a manner that guarantees that if a substring of a predefined length in one of the documents appears in another document, it will be detected, and thereby not rely on probability for measuring comparison accuracy." By referring to "substring of a

predefined length," it should be clear that Aiken does not concern itself with parts of speech. Further, "substring" indicates that there is no reordering." Continuing, at column 2, lines 51-57, Aiken states "In one aspect of the present invention, a method of comparing files and formatting output data involves receiving an input query file that can be segmented into multiple query file substrings. A query file substring is selected and used to search an index file containing multiple ordered file substrings that were taken from previously analyzed files." By referring to "ordered file substrings," it should be clear that Aiken is not concerned with reordering information and that Aiken is not concerned with parts of speech.

Aiken provides an example of the method disclosed therein beginning at column 14, line 18. This example has nothing to do with ordering and nothing to do with words. Fig. 3 confirms that Aiken is very different from the present invention. In Fig. 3, the phrase "this is it folks," is separated into substrings thi, is, isi, itf, fol and lks. The phrase is separated, but certainly not reordered. The separated portions have nothing to do with the meaning of the phrase.

None of the references relied upon by the Examiner taken alone or in any proper combination disclose or suggest reordering tokens of a document. Further, none of the references relied upon by the Examiner disclose or suggest filtering based on parts of speech. Therefore, the claims patentably distinguish over the references, and the prior art rejections should be withdrawn.

There being no further outstanding objections or rejections, it is submitted that the application is in condition for allowance. An early action to that effect is courteously solicited.

Finally, if there are any formal matters remaining after this response, the Examiner is requested to telephone the undersigned to attend to these matters.

If there are any additional fees associated with filing of this Amendment, please charge the same to our Deposit Account No. 19-3935.

<div style="text-align:center">Respectfully submitted,

STAAS & HALSEY LLP</div>

Date: Apr 13 2004

By: Mark J. Henry
Registration No. 36,162

1201 New York Avenue, NW, Suite 700
Washington, D.C. 20005
Telephone: (202) 434-1500
Facsimile: (202) 434-1501